Linguistic Approaches to Bilingualism

Less IS More: on the Tolerance Principle as a manifestation of Maximize Minimal Means

--Manuscript Draft--

Manuscript Number:	
Full Title:	Less IS More: on the Tolerance Principle as a manifestation of Maximize Minimal Means
Short Title:	Less IS More
Article Type:	Commentary
First Author:	Theresa Biberauer, PhD
Corresponding Author:	Theresa Biberauer, PhD Cambridge, UNITED KINGDOM
Funding Information:	
Section/Category:	
Keywords:	
Abstract:	
Author Comments:	This is the revised version of the paper that I submitted for consideration as a Response to/Commentary on Charles Yang's Target Article. I have shortened the ms, as requested (it now totals 1526 words, including 18 heading-related words), and also adjusted the presentation so that the components of the paper are more readily followable.

Less IS More: on the Tolerance Principle as a manifestation of Maximize Minimal Means

Theresa Biberauer

The formalist perspective on language acquisition Yang outlines may be 'personal' (Yang, 2018, p.1), but it offers a valuable illustration of the spirit guiding much 21st-century generative work. Central here is the drive to "reduce[] the explanatory burden traditionally placed on innate linguistic parameters" (Yang, 2018, p.5) by pursuing a three- (1a) rather than two-factor (1b) model of grammar acquisition and structure:

(1) a. Universal Grammar (UG) + Primary Linguistic Data (PLD) + general cognitive principles → an adult (L1) grammar

b. UG + PLD → an L1 grammar

Where (1b) foregrounded a richly specified UG, (1a) emphasizes a pared-down UG, and the grammar-shaping role of non-language-specific cognitive mechanisms, including "principles of data analysis … used in language acquisition and other domains", and "principles of efficient computation" (Chomsky, 2005, p.6). Yang's Tolerance Principle (TP) offers an explicit formal proposal as to how a potential third-factor principle - which rests on humans' well-established statistical sensitivity (Yang, 2004) - might interact with linguistic input to account for aspects of grammar acquisition and knowledge. As Yang convincingly demonstrates, the TP facilitates real insight into how "Less" can be "More" for child-acquirers. My purpose here is to endorse the "Less is More" spirit of Yang's model, and to suggest how the "periphery"-oriented TP may be

integrated with a similarly "Less is More"-oriented three-factors approach to "core" grammar, which potentially also sheds light on the difference between L1- and L2-acquirers.

The TP is strongly lexically oriented: the tolerance calculation requires the identification of a domain encompassing a specific number of lexical items to which a hypothesized generalization could apply. For the calculation to be meaningful - for the TP to drive acquisition - there additionally need to be exceptions to the hypothesized rule. These requirements allow us to situate the TP in relation to a recently proposed three-factors model, Biberauer (2017)'s so-called *Maximize Minimal Means (MMM)*-model.

The MMM-model postulates that child-acquirers aim to maximally exploit the knowledge at their disposal at all stages of the acquisition process, including the pre-lexical stage. MMM itself is conceived as a non-language-specific learning bias, which drives acquirers to construct their grammars incrementally on the basis of knowledge accessible to them at a given point. The idea is essentially that lesser access to input to begin with - less input which qualifies as **intake** - allows acquirers to focus their attention on a sub-component of the input, which they are then able to master and harness as the basis for access to more complex, previously inaccessible aspects of the input.

Much evidence suggests that L1-acquisition does indeed progress in this "Goldilocks" manner, with acquirers systematically attending to input that is neither too simple nor

too complex, but "just right" - thus parallelling what we see in other domains (e.g. vision; Kidd, Piatandosi & Aslin, 2012). Consider, for example, the research demonstrating *in utero* and very early post-birth sensitivity to prosody (Gervain & Werker, 2008 give an overview). Prosody delivers various 'edge'-oriented cues that allow acquirers to begin to "chunk" the input-strings in accordance with the grammar of their input-language(s) long before they have any lexical knowledge. Among other things, prosody alerts pre-lexical infants to basic head-directionality ("OV" = 'strong-weak' vs "VO" = 'weak-strong') and, during the first half-year, to the distinction between content and functional items, leaving 6-month-olds with a content-item preference. Thereafter, more fine-grained details become available, with, for example, the distribution of consonants and vowels within already-identified linguistic chunks facilitating the articulation of acquirers' knowledge of, respectively, vocabulary and inflectional morphology. Significant components of L1 knowledge are thus in place **before** the TP can regulate acquisition.

With some vocabulary in place, acquirers discover the key distinction between the truly arbitrary, memorization-requiring form-meaning mappings that characterize **content** items - classic Saussurean arbitrariness - and the "higher-level" arbitrariness that defines the recurring form and distribution of functional elements within a system - e.g. the obligatory inflection of Italian verbs, or the systematic fronting difference between lexical and auxiliary verbs in English questions. Biberauer (2017) proposes that such 'systematic departures from Saussurean arbitrariness' cue the postulation of grammar-defining formal ([F])-features alongside content-item-defining semantic and phonological features (Chomsky, 1995). Two points are crucial: firstly, [F]s define the

formal make-up of functional categories, and, hence, parametric variation, or "core grammar". Secondly, the postulation of [F]s, in accordance with MMM-driven Feature Economy ('Postulate as few [F]s as possible to account for observed regularities') and Input Generalization ('Maximize the use of already-postulated [F]s'), allows the acquirer to optimize their knowledge of system-defining regularities in a manner parallelling that driving the TP-governed postulation of "periphery"-oriented rules: both reflect a response to the acquirer's 'search for [memorization-limiting - TB] productive generalizations' (Yang 2018, p.14), which is, at base, driven by MMM. Further, to the extent that [F]s have both grammar- (Narrow Syntax) internal and realizational (PF) consequences - triggering Agree and Move operations, and determining placement and morphological realization - we can begin to refine our understanding of how "core" [F]-mediated regularities and TP-sanctioned peripheral rules relate to one another, and what acquisitional consequences one might expect.

Significantly, the conclusion that peripheral rules are more "surfacey", with the TP merely regulating realizational options, seems too simple. Consider the "size"-based emergent (non-UG-given) parameter typology (2), and the associated 'learning pathway' (3):

(2) For a given value v_i of a parametrically variant feature [F]:

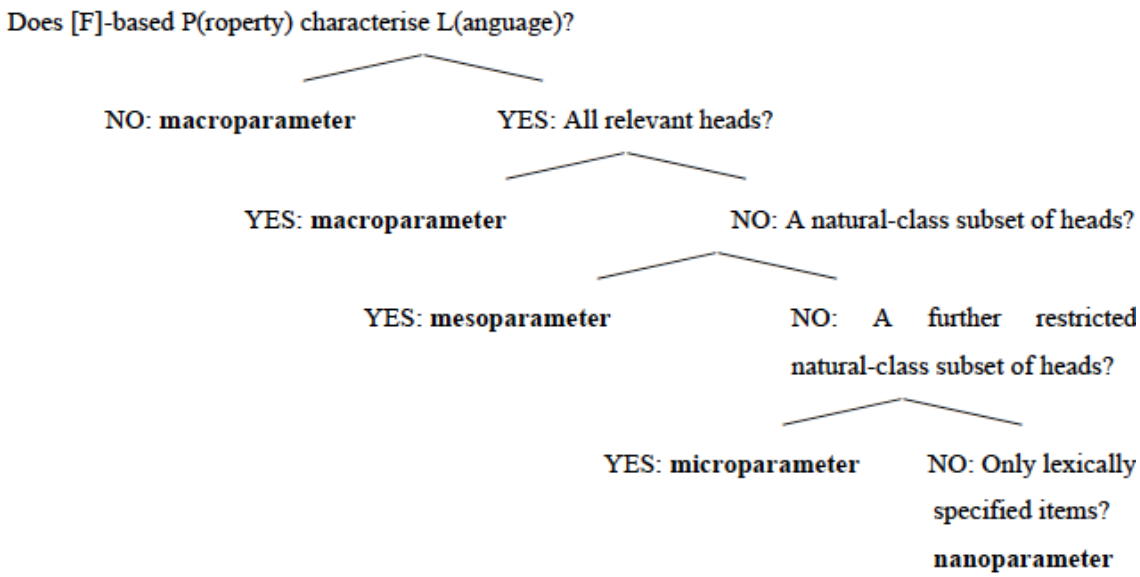
a. **Macroparameters**: All heads of the relevant type, e.g. all probes/phase heads, share v_i. Examples: harmonic head-finality, radical pro-drop.

b. **Mesoparameters**: All heads in a given natural class, e.g. [+V] or all C-/T-heads, share v_i. Examples: Chinese-style head-finality, Romance-style pro-drop, Verb-Second.

c. **Microparameters**: A small, lexically definable sub-class of functional heads, e.g. (modal) auxiliaries, subject clitics, show v_i. Examples: partial pro-drop, English-style Verb-Second.

d. **Nanoparameters**: One/more individual lexical items is/are specified for v_i. Examples: Bavarian-style pro-drop, English-style Conditional Inversion.

(3)



Key here is the idea that macro-, meso- and microparameters are formulated over natural classes, namely, categories defined over more/less detailed [F]-specifications. L1-acquisition is largely assumed to proceed "top-down", with the acquirer's initially underspecified grammar progressively becoming more articulated in [F] terms as more of the input becomes accessible, and earlier acquisition stages, accordingly, featuring various underspecified "shadow" phenomena. More restricted/specialized natural classes are featurally more complex, meaning that we expect "smaller" [F]-regulated

phenomena to be fully acquired later than "bigger" ones. Importantly, "bigger" class-based parameters do not seem to require acquirers to attend to TP-relevant lexical (ir)regularities.

Nanoparameters are different, however: they govern non-[F]-unified individual lexical items, are acquired bottom-up, independently of the general top-down process, and may therefore be acquired early or late, depending on the complexity of the [F]s encoded on the nanoparametrically specified items. Because nanoparameters don't target an entire natural class - consider (standard) English Conditional Inversion (CI), which affects only a subset of English's seven past-marked auxiliaries: *had*, *were*, *should* - and because the class is quantifiable, they seem to exhibit the same profile as TP-sanctioned "peripheral" rules. And English CI in fact satisfies the TP (N/lnN = 1.946). In principle, given the definition of nanoparameter, this need not always be the case, however, leading us -seemingly correctly - to expect TP-regulated instability in the nanoparametric corners of grammar. The TP may thus regulate not just the productivity of regular rules like those discussed in Yang's work, but also the stability of certain types of exceptions, the nanoparameters, which often instantiate "left-overs" from a more productive system.

What does this imply for our understanding of the differences between L1- and (adult) L2-acquisition? Yang suggests that adult learners' larger vocabularies effectively undermine their ability to harness the TP with an L1-acquirer's effectiveness, leaving them unable to reap "Less is More"-type benefits. From an MMM perspective, this may be part of the problem; more generally, however, it seems that adults' abundance of

highly-developed resources - language-specific and general-cognitive - compromises their ability to "get back to basics" so as to build up their grammars incrementally, via the "Goldilocks"-route taken by L1-acquirers. If our discussion is on the right track, it will not just be the lexically-based and thus potentially TP-regulated aspects of grammar that L2-acquirers will achieve variable success with; invariant and thus non-TP-regulated "bigger" properties will also be undercut by the knowledge and biases that adults bring to the task. And, if that is correct, three-factors-inspired formal work has much to offer to deepen our understanding of the surprising properties of L1-acquisition, mono- and bilingual, and of the nature of the challenge facing L2-learners.

References

Biberauer, T. (2017). Factors 2 and 3: a principled approach. *Cambridge Occasional Papers in Linguistics* 10: 38-65.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry 36*, 1-22.

Gervain, J. & Werker, J. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass 2*(6), 1149-1170.

Kidd, C., Piatandosi, S. & Aslin, R. (2012). The Goldilocks Effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE 7*(5), e36399.

Yang, C. (2004). Universal Grammar, statistics, or both? *Trends in Cognitive Science 10*, 451-456.

Author's address

Department of Computer Science and Technology

The Computer Laboratory

William Gates Building

15 JJ Thompson Avenue

University of Cambridge

Cambridge

CB3 0FD

E-mail: mtb23@cam.ac.uk